FIG. 1

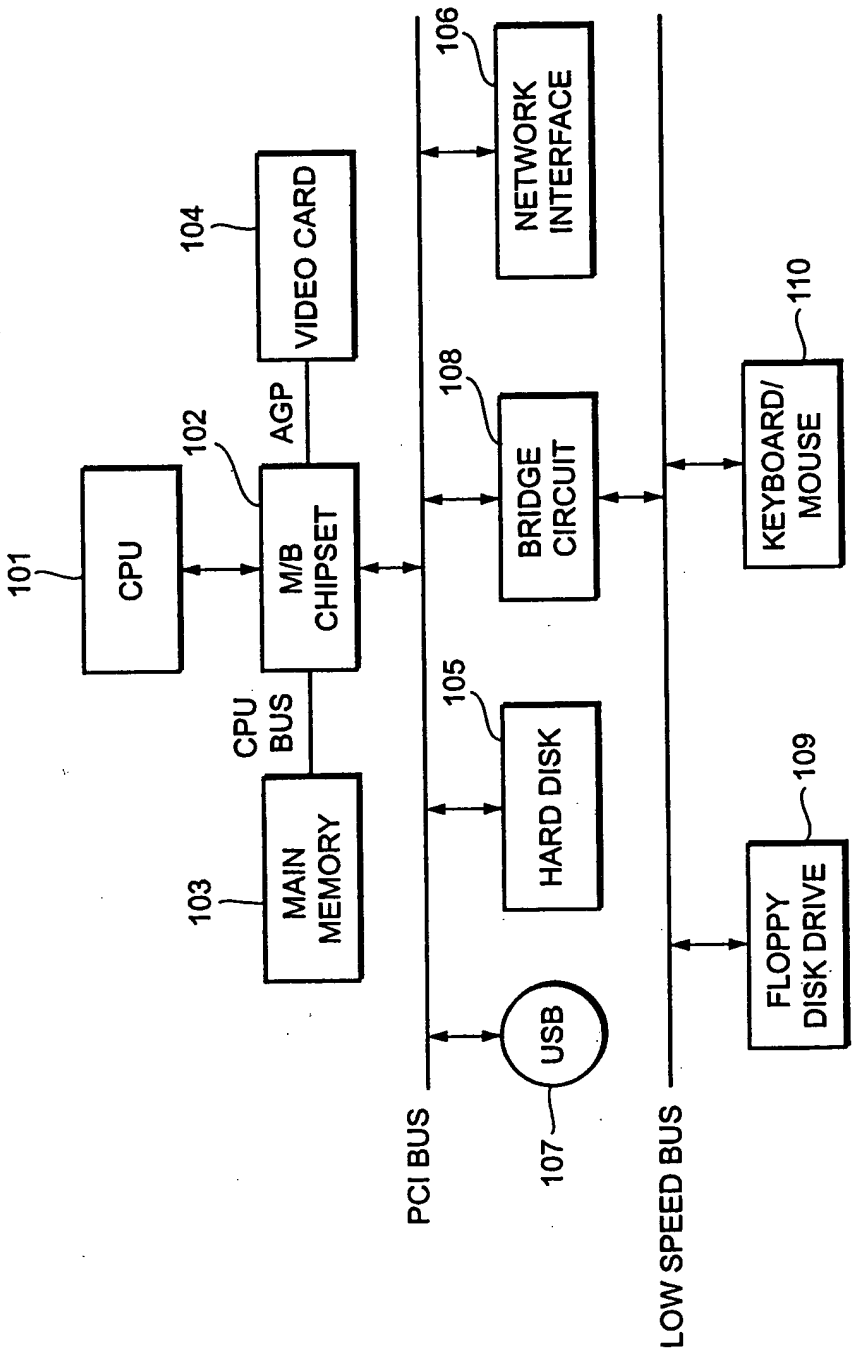
101
CPU
102
M/B CHIPSET
AGP
104
VIDEO CARD
CPU BUS
103
MAIN MEMORY
106
NETWORK INTERFACE
108
BRIDGE CIRCUIT
105
HARD DISK
107
USB
PCI BUS
110
KEYBOARD/ MOUSE
109
FLOPPY DISK DRIVE
LOW SPEED BUS

FIG. 2

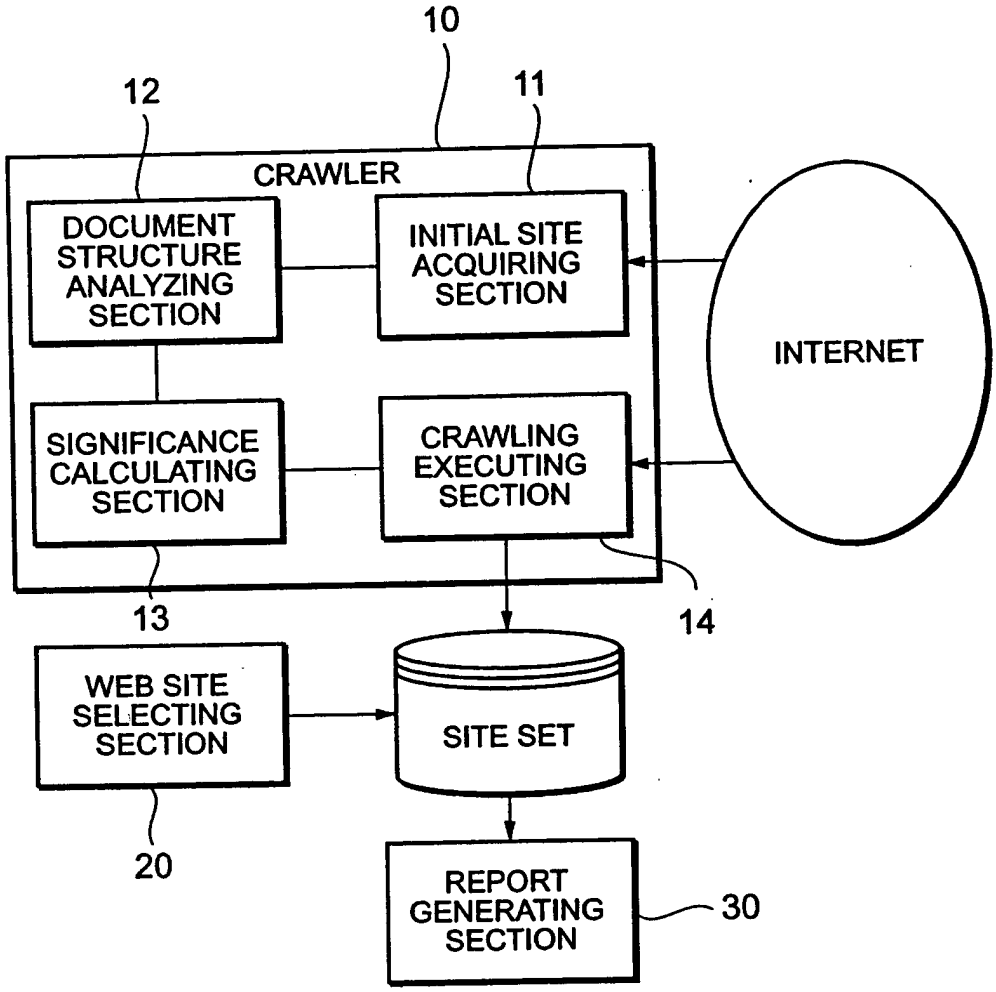
10
12
11
CRAWLER
DOCUMENT STRUCTURE ANALYZING SECTION
INITIAL SITE ACQUIRING SECTION
INTERNET
SIGNIFICANCE CALCULATING SECTION
CRAWLING EXECUTING SECTION
13
14
WEB SITE SELECTING SECTION
SITE SET
20
REPORT GENERATING SECTION
30

FIG. 3

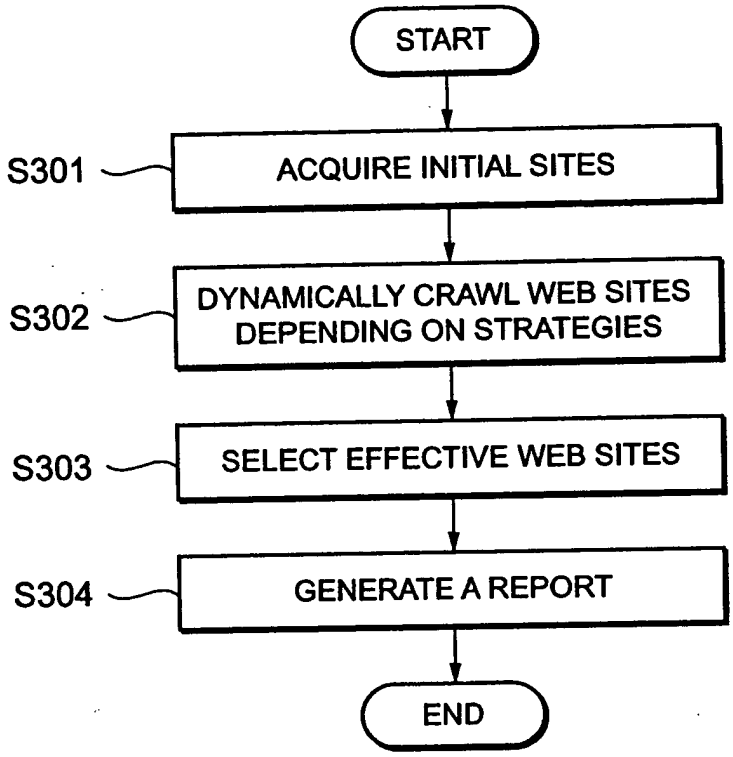
START
S301
ACQUIRE INITIAL SITES
S302
DYNAMICALLY CRAWL WEB SITES DEPENDING ON STRATEGIES
S303
SELECT EFFECTIVE WEB SITES
S304
GENERATE A REPORT
END

FIG. 4

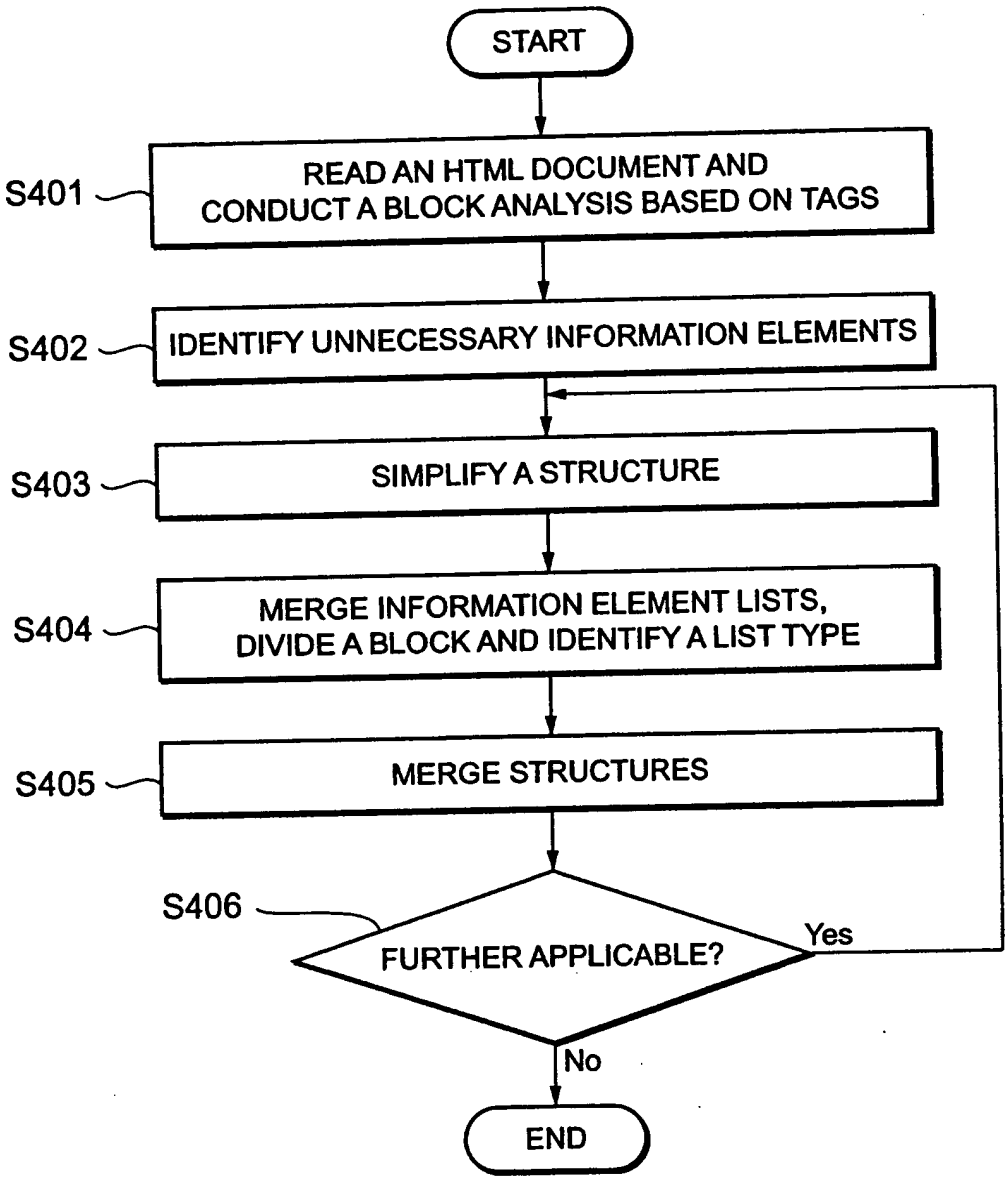
START
S401
READ AN HTML DOCUMENT AND CONDUCT A BLOCK ANALYSIS BASED ON TAGS
S402
IDENTIFY UNNECESSARY INFORMATION ELEMENTS
S403
SIMPLIFY A STRUCTURE
S404
MERGE INFORMATION ELEMENT LISTS, DIVIDE A BLOCK AND IDENTIFY A LIST TYPE
S405
MERGE STRUCTURES
S406
FURTHER APPLICABLE?
Yes
No
END

FIG. 5A

```
⋮
<img src="bullet.gif">ITEM 1<br>
<img src="bullet.gif">ITEM 2<br>
⋮
```

FIG. 5B

```
⋮
<a href="top.htm">GO TO TOP</a>
⋮
<a href="top.htm">GO TO TOP</a>
```

FIG. 6

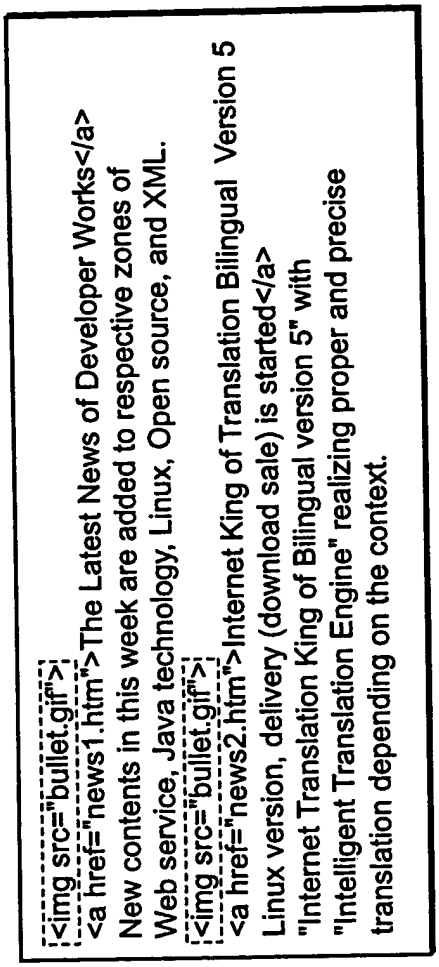
<img src="bullet.gif">
The Latest News of Developer Works
New contents in this week are added to respective zones of
Web service, Java technology, Linux, Open source, and XML.
<img src="bullet.gif">
Internet King of Translation Bilingual Version 5
Linux version, delivery (download sale) is started
"Internet Translation King of Bilingual version 5" with
"Intelligent Translation Engine" realizing proper and precise
translation depending on the context.

FIG. 7

```
<img src="bullet.gif">
<a href="news1.htm">The Latest News of Developer Works</a>
New contents in this week are added to respective zones of
Web services, Java Technology, Linux, Open source, and XML.
for details, click <a href="detail.htm">have </a>
<img src="bullet.gif">
<a href="news2.htm">Internet Translation King of Bilingual Version 5
Linux version, delivery (download sale) is started</a>
"Internet Translation King of Bilingual Version 5" with
"Intelligent Translation Engine" realizing proper and precise
translation depending on the context.
```

FIG. 8

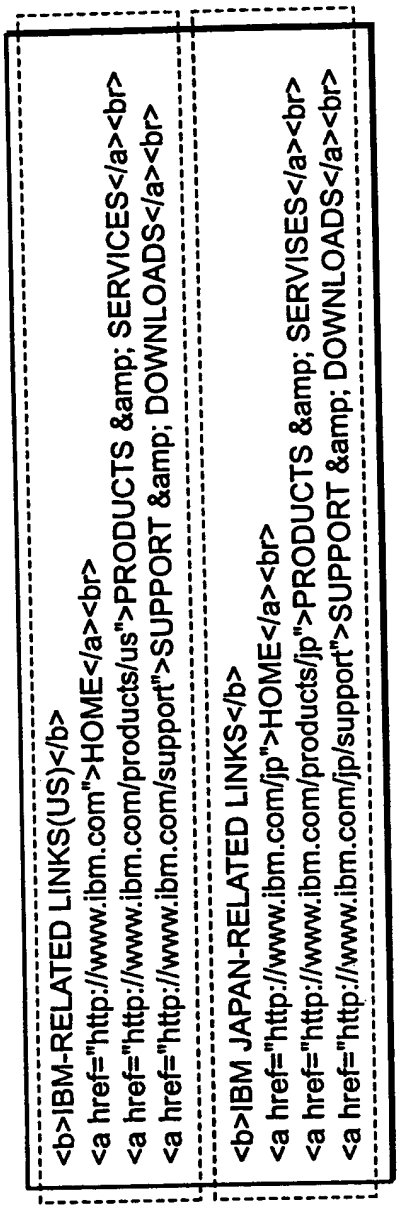
IBM-RELATED LINKS(US)
HOME
PRODUCTS & SERVICES
SUPPORT & DOWNLOADS
IBM JAPAN-RELATED LINKS
HOME
PRODUCTS & SERVISES
SUPPORT & DOWNLOADS

FIG. 9

```
<block>
<block title="IBM-RELATED LINKS(US)">
 <object type="anchor" url="http://www.ibm.com" title="HOME"/>
 <object type="anchor" url="http://www.ibm.com/products/us" title="PRODUCTS &amp;amp; SERVICES"/>

 <object type="anchor" url="http://www.ibm.com/support" title="SUPPORT &amp;amp; DOWNLOADS"/>
 </block>
<block title="IBM JAPAN-RELATED LINKS">
 <object type="anchor" url="http://www.ibm.com/jp" title="HOME"/>
 <object type="anchor" url="http://www.ibm.com/products/jp" title="PRODUCTS &amp;amp; SERVICES"/>

 <object type="anchor" url="http://www.ibm.com/jp/support" title="SUPPORT &amp;amp; DOWNLOADS"/>
 </block>
 </block>
```

FIG. 10

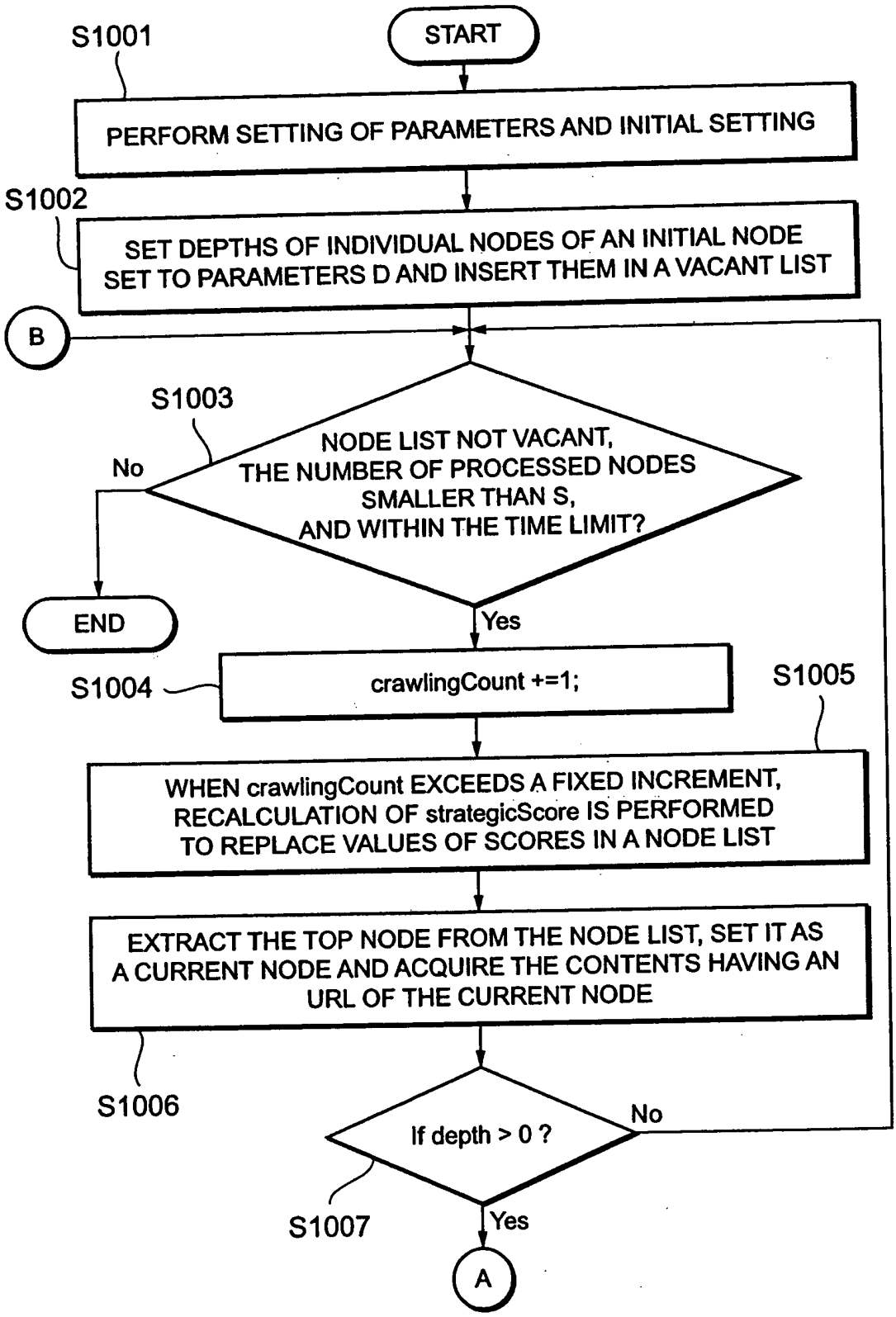
S1001
START
PERFORM SETTING OF PARAMETERS AND INITIAL SETTING
S1002
SET DEPTHS OF INDIVIDUAL NODES OF AN INITIAL NODE SET TO PARAMETERS D AND INSERT THEM IN A VACANT LIST
B
S1003
NODE LIST NOT VACANT, THE NUMBER OF PROCESSED NODES SMALLER THAN S, AND WITHIN THE TIME LIMIT?
No
END
Yes
S1004
crawlingCount +=1;
S1005
WHEN crawlingCount EXCEEDS A FIXED INCREMENT, RECALCULATION OF strategicScore IS PERFORMED TO REPLACE VALUES OF SCORES IN A NODE LIST
EXTRACT THE TOP NODE FROM THE NODE LIST, SET IT AS A CURRENT NODE AND ACQUIRE THE CONTENTS HAVING AN URL OF THE CURRENT NODE
S1006
If depth > 0 ?
No
S1007
Yes
A

FIG. 11

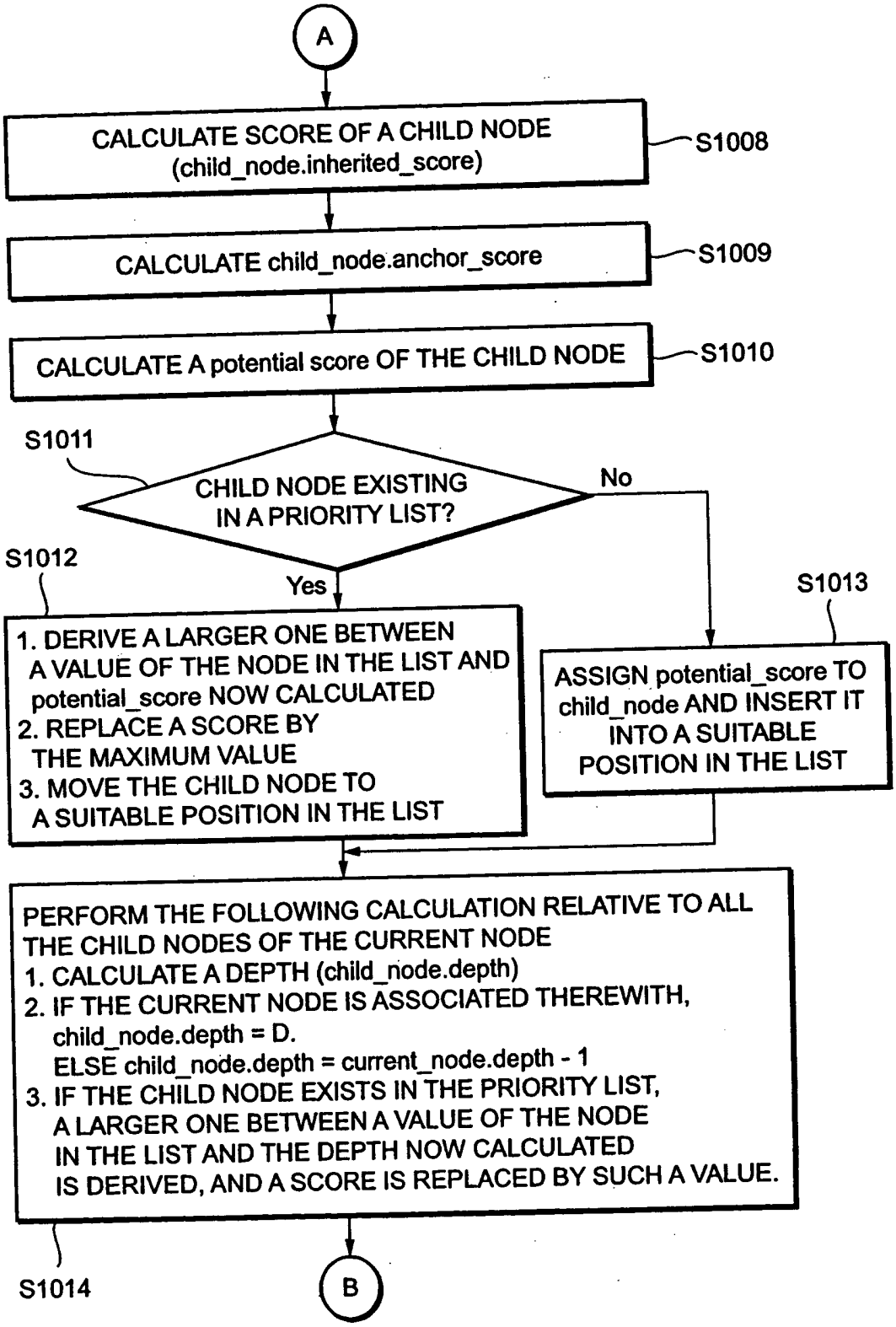
A
CALCULATE SCORE OF A CHILD NODE (child_node.inherited_score)
S1008
CALCULATE child_node.anchor_score
S1009
CALCULATE A potential score OF THE CHILD NODE
S1010
S1011
CHILD NODE EXISTING IN A PRIORITY LIST?
No
Yes
S1012
1. DERIVE A LARGER ONE BETWEEN A VALUE OF THE NODE IN THE LIST AND potential_score NOW CALCULATED
2. REPLACE A SCORE BY THE MAXIMUM VALUE
3. MOVE THE CHILD NODE TO A SUITABLE POSITION IN THE LIST
S1013
ASSIGN potential_score TO child_node AND INSERT IT INTO A SUITABLE POSITION IN THE LIST
PERFORM THE FOLLOWING CALCULATION RELATIVE TO ALL THE CHILD NODES OF THE CURRENT NODE
1. CALCULATE A DEPTH (child_node.depth)
2. IF THE CURRENT NODE IS ASSOCIATED THEREWITH, child_node.depth = D.
ELSE child_node.depth = current_node.depth - 1
3. IF THE CHILD NODE EXISTS IN THE PRIORITY LIST, A LARGER ONE BETWEEN A VALUE OF THE NODE IN THE LIST AND THE DEPTH NOW CALCULATED IS DERIVED, AND A SCORE IS REPLACED BY SUCH A VALUE.
S1014
B

FIG. 12

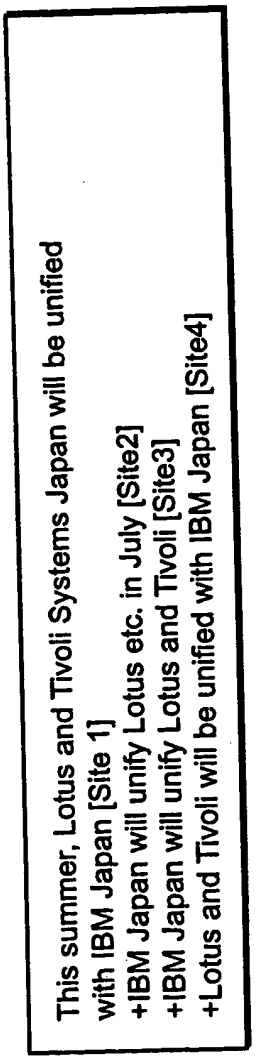
This summer, Lotus and Tivoli Systems Japan will be unified with IBM Japan [Site 1]
+IBM Japan will unify Lotus etc. in July [Site2]
+IBM Japan will unify Lotus and Tivoli [Site3]
+Lotus and Tivoli will be unified with IBM Japan [Site4]

FIG. 13

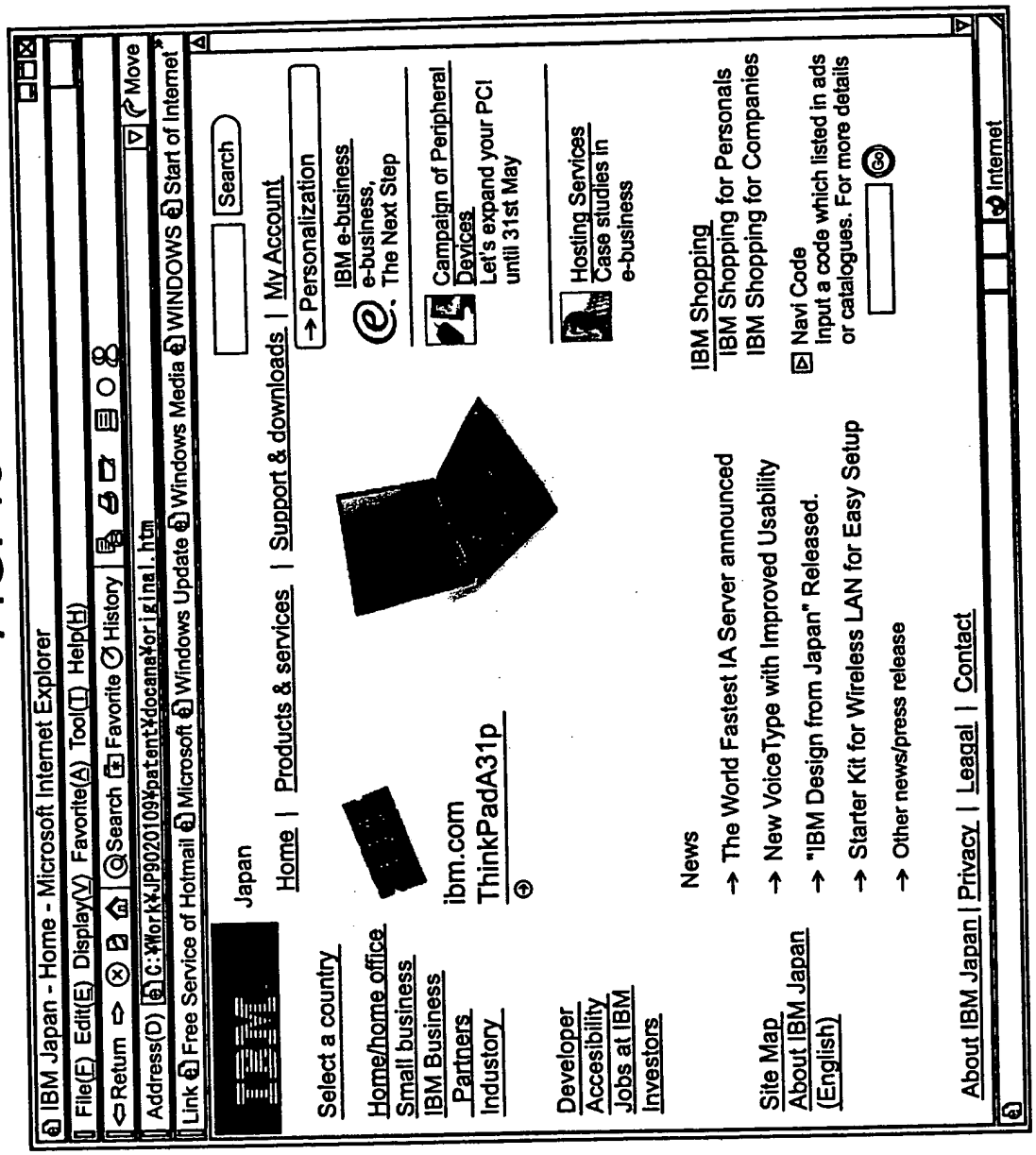
IBM Japan - Home - Microsoft Internet Explorer
File(F) Edit(E) Display(V) Favorite(A) Tool(T) Help(H)
Return Search Favorite History Move
Address(D) C:¥Work¥JP9020109¥patent¥docana¥original.htm
Link Free Service of Hotmail Microsoft Windows Update Windows Media WINDOWS Start of Internet
Japan
Search
Home | Products & services | Support & downloads | My Account
Personalization
Select a country
Home/home office
Small business
IBM Business Partners
Industory
ibm.com
ThinkPadA31p
IBM e-business
e-business, The Next Step
Campaign of Peripheral Devices
Let's expand your PCI until 31st May
Hosting Services
Case studies in e-business
Developer
Accesibility
Jobs at IBM
Investors
News
→ The World Fastest IA Server announced
→ New VoiceType with Improved Usability
→ "IBM Design from Japan" Released.
→ Starter Kit for Wireless LAN for Easy Setup
→ Other news/press release
IBM Shopping
IBM Shopping for Personals
IBM Shopping for Companies
Navi Code
Input a code which listed in ads or catalogues. For more details
Go
Site Map
About IBM Japan
(English)
About IBM Japan | Privacy | Leagal | Contact
Internet

FIG. 14

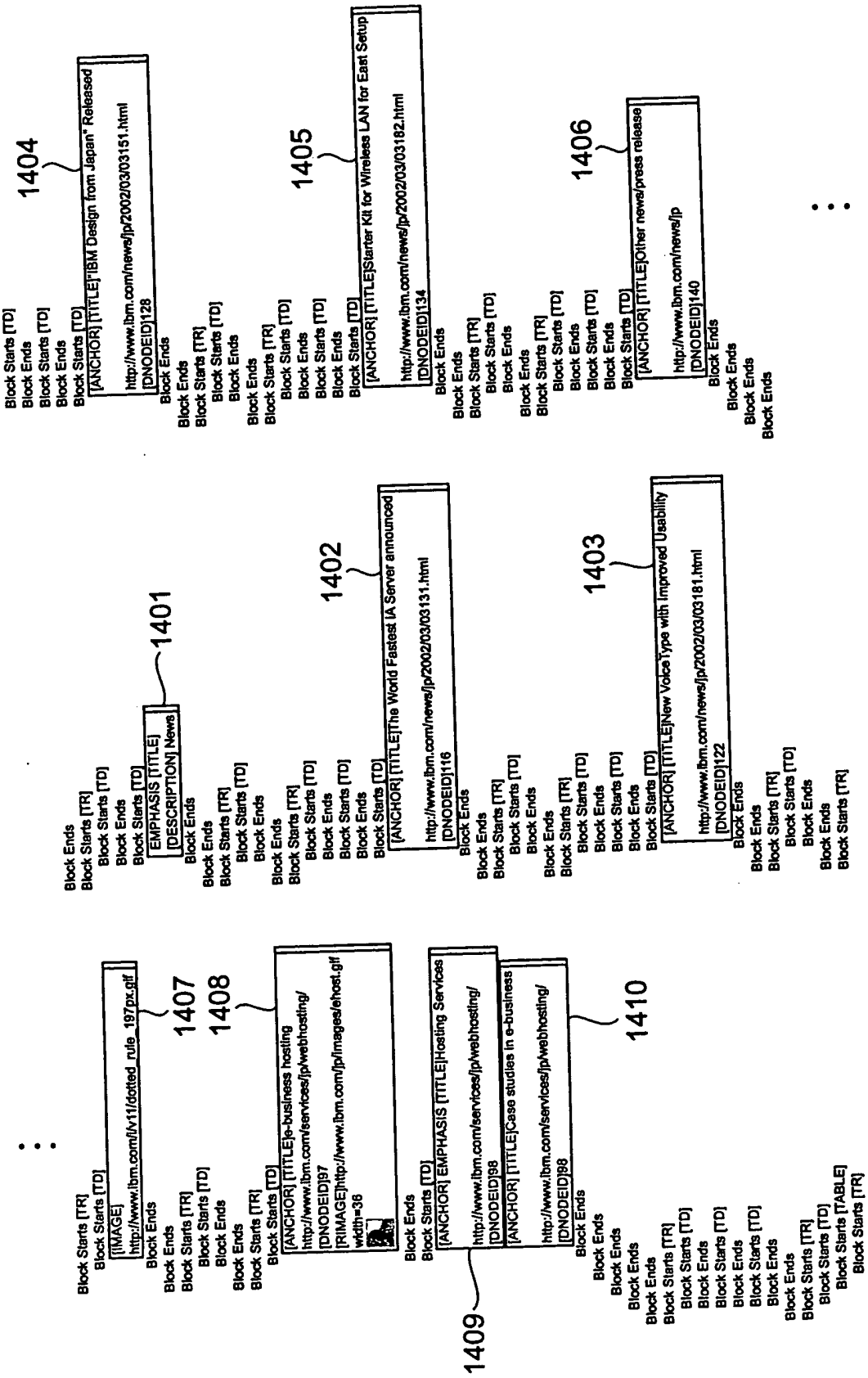
Block Starts [TR]
Block Starts [TD]
[IMAGE]
http://www.ibm.com/i/v11/dotted_rule_197px.gif
1407
1408
[ANCHOR] [TITLE]e-business hosting
http://www.ibm.com/services/jp/webhosting/
[DNODEID]97
[IMAGE]http://www.ibm.com/jp/images/ehost.gif
width=36
1409
[ANCHOR] EMPHASIS [TITLE]Hosting Services
http://www.ibm.com/services/jp/webhosting/
[DNODEID]98
[ANCHOR] [TITLE]Case studies in e-business
http://www.ibm.com/services/jp/webhosting/
[DNODEID]98
1410
[EMPHASIS [TITLE]
[DESCRIPTION] News
1401
[ANCHOR] [TITLE]The World Fastest IA Server announced
http://www.ibm.com/news/jp/2002/03/03131.html
[DNODEID]116
1402
[ANCHOR] [TITLE]New VoiceType with Improved Usability
http://www.ibm.com/news/jp/2002/03/03181.html
[DNODEID]122
1403
[ANCHOR] [TITLE]"IBM Design from Japan" Released
http://www.ibm.com/news/jp/2002/03/03151.html
[DNODEID]128
1404
[ANCHOR] [TITLE]Starter Kit for Wireless LAN for East Setup
http://www.ibm.com/news/jp/2002/03/03182.html
[DNODEID]134
1405
[ANCHOR] [TITLE]Other news/press release
http://www.ibm.com/news/jp
[DNODEID]140
1406

FIG. 15

Block Starts [TD] TYPE=SITE_MAP

[ANCHOR] [TITLE]Personalization http://www.ibm.com/cgi-bin/ipc/prefs/jp [DNODEID]15 [RIMAGE]http://www.ibm.com/i/prefs/v11/homepage_site_pref_tab_bott.gif width=198	personalization[13]
[ANCHOR] [TITLE]e-business logo:IBM e-business http://www.ibm.com/jp/e-business/ [DNODEID]15 [DESCRIPTION]e-business, The Next Step [RIMAGE]http://www.ibm.com/jp/images/img_ebusiness.gif width=36	e-business[13] logo[TO]
[ANCHOR] [TITLE]spring option campaign:Campaign of Peripheral Devices http://www.ibm.com/shop/jp [DNODEID]15 [DESCRIPTION]Let's expand your PC! until 31st May [RIMAGE]http://www.ibm.com/jp/images/spring_campaign.gif width=36	campaign[11] option[13] spring[11]
[ANCHOR] [TITLE]e-business hosting:Hosting Services http://www.ibm.com/services/jp/webhosting/ [DNODEID]15 [DESCRIPTION]Case studies in e-business [RIMAGE]http://www.ibm.com/jp/images/ehost.gif width=36	e-business[13] host[12]

Block Ends

1502

1501

Block Ends

Block Starts [TR]

Block Starts [TD] TYPE=TEXTPARAGRAPHTITLE=[News]

EMPHASIS [TITLE] [DESCRIPTION] News	news[13]
[ANCHOR] [TITLE]The World Fastest IA Server announced http://www.ibm.com/news/jp/2002/03/03131.html [DNODEID]17	IA[TO] server[14] fast[AD] world[12] announce[13]
[ANCHOR] [TITLE]New VoiceType with Improved Usability http://www.ibm.com/news/jp/2002/03/03181.html [DNODEID]17	VoiceType[14] improve[11] new[AD] usability [11] announce[13]
[ANCHOR] [TITLE]"IBM Design from Japan" Released http://www.ibm.com/news/jp/2002/03/03151.html [DNODEID]17	Design[TO] IBM[OR] Japan[CO] from[TO] release[13]
[ANCHOR] [TITLE]Starter Kit for Wireless LAN for Easy Setup http://www.ibm.com/news/jp/2002/03/03182.html [DNODEID]17	starter kit[TO] setup[13] easy[AD] wireless LAN[14]
[ANCHOR] [TITLE]Other news/press release http://www.ibm.com/news/jp [DNODEID]17	news[13] press release[13]

Block Ends